

明 細 書

発明の名称：

変異プロファイリングのためのRNAプローブ及びその使用

クロスリファレンス

[0001] 本出願は、日本国において、2020年11月18日に出願された特願2020-191550号に基づく優先権を主張するものであり、当該出願に記載された内容は全て、参照によりそのまま本明細書に援用される。また、本願において引用した全ての特許、特許出願及び文献に記載された内容は全て、参照によりそのまま本明細書に援用される。

技術分野

[0002] 本発明は、変異プロファイリングのためのRNAプローブに関し、より詳細には、解析対象RNAに構造化バーコード配列を付加したRNAプローブ及びそれを用いてRNAの高次構造を解析する方法に関する。

背景技術

[0003] RNAはタンパク質合成の鋳型として機能する生体分子であるが、一方で、RNA自身が密に折りたたまれた高次構造を形成し、遺伝子発現、転写産物の細胞内局在及びスプライシング機構などを制御している。これらの機能性RNAの多くは、一次配列としての塩基が構造形成にて立体的に特定の配置をとることで規定されている。このRNA高次構造は、ステム（STEM）、ステムループ（STEM-LOOP）、キッシングループ（KISSING-LOOP）、マルチジャンクション（MULTI-JUNCTION）、キンクターン（KINK-TURN）、シュードノット（PSEUDO KNOT）、4重鎖（QUADRUPLEX）などの多様な構造モチーフの組み合わせから形成される。これらの構造モチーフの種類と組み合わせの数は膨大である上に、複数の平衡状態をとり得るため予測が困難である。これに対し、特定の塩基に対する化学的な修飾反応と、次世代シーケンシングによって得られた配列データを組み合わせてRNA高次構造を決定する技術が

開発されている。

[0004] 例えば、選択的に核酸の糖の2位の炭素に修飾を加えるSHAPE-MaP法（特許文献1参照）や、硫酸ジメチル（DMS）を用いるDMS-MaPseq法（非特許文献1参照）などの変異プロファイリング（Mutational Profiling: MaP）が、RNAの二次構造を推定するために使用されている。化学修飾の分布はRNAの二次構造と相関しており、相補DNAの塩基配列を決定する際に、位置特異的な逆転写終結、置換、挿入又は欠失による変異として記録される。

[0005] 変異プロファイリングは次世代シーケンシングとの統合によって、より広範な種類のRNAを同時解析可能である。例えばDMS-MaPseq法やSHAPE-MaP法においては、変異が導入されたRNA由来のDNA断片をコンピュータ上にてリファレンスゲノムに対してマッピングする。この操作によって複数種類が混合された条件において配列が仕分けされ、複数領域や異なる分子のRNAに対して構造特異的な変異を同時カウントできる。またPORE-cupine法などにおいてはナノポアシーケンサーによって修飾種が与える直接的な電位の変化の検出により、同一分子内の変異を複数種類カウントできる（例えば、非特許文献2参照）。しかしながら、これらは配列の仕分けにリファレンスゲノムに対するマッピング操作を使用しているため、類似する配列が存在している場合にどのゲノム位置由来かどうかわからなくなる欠点が存在する。例えば、遺伝子ファミリー、アリル特異的なRNAなどがあげられる。さらに、RNA修飾試薬による変異導入は類似する配列の多様性を増加させるため、この効果を増長する。

先行技術文献

非特許文献

[0006] 非特許文献1: Megan Zubradt et al. DMS-Mapsseq for genome-wide or targeted RNA structure probing in vivo. Nat Methods. 14, 75-82 (2017)

非特許文献2: Aw, J. G. A. , Lim, S. W. , Wang, J. X. et al. Determination of isoform-specific RNA structure with nanopore long reads. Nat Biotechnol (2020). <https://doi.org/10.1038/s41587-020-0712-z>

特許文献

[0007] 特許文献1: 特許第6612220号公報

発明の概要

発明が解決しようとする課題

[0008] 多種類のRNAを含むRNAライブラリを用いて上述した化学修飾による変異プロファイリングを行う場合、1つ又は数個の塩基のみが異なる配列からなるRNAについては、化学修飾により生じた変異と、ライブラリ中に含まれ得る天然の多様な配列とを相補DNAの整列により識別することが難しいという問題があった。

[0009] 本発明は、RNAライブラリを用いて変異プロファイリングを行う場合に、解析対象のRNA高次構造に影響を与えることなく、導入された塩基の変異、挿入及び欠失等の検出精度を改善することを課題とする。

課題を解決するための手段

[0010] 本発明は、かかる課題を解決するためになされたものであって、変異プロファイリングを行う際に、RNAライブラリに含まれるそれぞれのRNAに、異なる固有の配列であって化学修飾剤との反応が抑制されるような構造を有するバーコード配列を付加するようにした。

[0011] すなわち、本発明の第一の視点において、RNAの高次構造を解析するための方法は、(a) 解析対象RNAにバーコード配列を付加した1又は複数のRNAプローブを調製する工程、(b) RNAプローブとRNA修飾剤とを接触させる工程、及び(c) 工程(b)で得られたRNAプローブの配列

中で修飾を受けた塩基の位置と頻度を検出する工程、を含む。このバーコード配列はRNA修飾剤との反応が抑制される構造を有することと解析対象RNAと高次構造を形成しないことを特徴とする。上記検出工程（c）は、以下の工程を含むことが好ましい。

（c1）工程（b）で得られたRNAプローブの混合物を鋳型として逆転写酵素により相補DNAを合成する工程、（c2）相補DNAの塩基配列を決定し、バーコード配列を含む塩基配列を整列させる工程、及び（c3）整列させた塩基配列に生じた変異の位置と頻度を検出する工程。

[0012] 本発明の他の視点では、複数の塩基対を含む構造を形成するバーコード配列が付加された解析対象RNAを含むRNAプローブ及び、このRNAプローブの複数を含むRNAプローブライブラリが提供される。さらなる実施形態では、このRNAプローブライブラリの2以上の複製物からなるRNAプローブライブラリ群が提供される。複製されたすべてのRNAプローブは、さらに第2のバーコード配列を含み、この第2のバーコード配列は、1つのライブラリ内ではすべて同一配列であるが他のライブラリとの間では識別可能である。

発明の効果

[0013] 本発明によれば、RNAライブラリを用いて変異プロファイリングを行う場合に、解析対象のRNA高次構造に影響を与えることなく、導入された塩基の変異、挿入及び欠失等の検出精度を改善することができる。

図面の簡単な説明

[0014] [図1]図1は、一実施形態におけるRNAの高次構造の解析方法を示すフロー図である。

[図2]図2は、別の実施形態におけるRNAの高次構造の解析方法を示すフロー図である。

[図3]図3は、第1のライブラリの作製に用いたバーコード配列（a）及びライブラリ構造の概要（b）を表した模式図である。

[図4]図4は、37種類の第1のバーコード配列と、4種類の第2のバーコー

ド配列(バッチバーコード)を用いて作製したライブラリ構造の概要を表した模式図である。

[図5]図5は、第1のライブラリに含まれるRNAプローブのうち、個別のストランドとして合成した2つのサンプル(ID1及びID32)の塩基配列である。

[図6]図6は、第2のライブラリを用いて行った変異プロファイリング操作の流れを示す模式図である。

[図7]図7は、NAI又はDMSで化学修飾されたサンプルのバーコード内のすべてのヌクレオチドのデルタ変異率の絶対値を示す。結果は、第1のライブラリにおけるRNAプローブ中の構造化バーコード(ID1-28)と非構造化バーコード(ID29-37)に分けて示した。

[図8]図8は、各ライブラリをNAI又はDMSで化学修飾したときの、各ヌクレオチドについてのデルタ変異率を示した結果である。X軸は、ID1の目的RNAの配列及びドット・ブラケット表記法による推定構造を示す。(a)は、第1のライブラリ及び4種類の第2のライブラリをNAIで処理したときの結果であり、(b)は、第1のライブラリ及び4種類の第2のライブラリをDMSで処理したときの結果であり、(c)は、第2のライブラリをそれぞれ個別に又はプールしてNAIで処理したときの結果であり、(d)は、第2のライブラリをそれぞれ単独で又はプールしてDMSで処理したときの結果である。

[図9]図9は、第2のライブラリをそれぞれ個別に又はプールしてNAI又はDMSで化学修飾したときの、各IDのデルタ変異率を、塩基対を形成する領域(黒い部分)及び形成しない領域(灰色の部分)と予測されるヌクレオチドのデルタ変異率のカーネル密度分布を示すバイオリンプロットである。

(a)は、NAIで処理したサンプル、(b)は、DMSで処理したサンプルである。それぞれの「バイオリン」の左側は、一緒に取得された4つの個別のサンプルの分布、右側はプールされたサンプルの分布である。DMSで処理したサンプルの場合、数値計算ではCとAのみの変異を考慮した。

[図10]図10は、構造化バッチバーコードを付加したRNAプローブライブラリ群を用いて修飾剤なしで変異プロファイリングを行い、ID1が付与されたRNAプローブライブラリの次世代シーケンスで得られたすべてのリードを、ID1から96のバーコードが付与されたRNAプローブライブラリ群のファイルについてマッピングしたときのリード数を各IDについてプロットしたグラフである。

[図11]図11は、構造化バッチバーコードを付加したRNAプローブライブラリ群を用いてDMSによる変異プロファイリングを行い、ID2が付与されたRNAプローブライブラリの次世代シーケンスで得られたすべてのリードを、ID1から96のバーコードが付与されたRNAプローブライブラリ群のファイルについてマッピングしたときのリード数を各IDについてプロットしたグラフである。

[図12]図12は、構造化バッチバーコードを付加したRNAプローブライブラリ群を用いて修飾剤なしで変異プロファイリングを行い、夫々のRNAごとに正しいIDと判定された割合をプロットした結果を示す。

[図13]図13は、構造化バッチバーコードを付加したRNAプローブライブラリ群を用いてDMSによる変異プロファイルリングを行い、夫々のRNAごとに正しいIDと判定された割合をプロットした結果を示す。

[図14]図14は、構造化バッチバーコードを用いて変異プロファイリングを行った後、複数のインデックスと組み合わせて次世代シーケンシングを行い、ID7が付与されたRNAプローブライブラリから得られたすべてのリードを、ID1から96のバーコードが付与されたRNAプローブライブラリ群についてマッピングしたときのリード数を各IDについてプロットしたグラフである。

[図15]図15は、構造化バッチバーコードに1対1で対応するインデックスを付与して行った次世代シーケンシングの結果、インデックスIDに対してマッピングされた構造化バッチバーコードIDのリード数をプロットしたグラフである。

[図16]図16は、図15において、各構造化バッチバーコードIDが付与されたRNAプローブライブラリの中で誤判定されたRNAの種類(RNA ID)数をプロットしたグラフである。

[図17]図17は、構造化バッチバーコードに1対1で対応するインデックスを付与して次世代シーケンシングを行い、構造化バッチバーコードのIDの判定における正確性をインデックスごとにプロットした結果である。

[図18]図18は、実施例4で用いた構造化バッチバーコード配列の例(ID 12及びID 28)を示す図である。

発明を実施するための形態

[0015] 次に、本発明の各実施形態について、図面を参照して説明する。なお、以下に説明する各実施形態は、特許請求の範囲に係る発明を限定するものではなく、また、各実施形態の中で説明されている諸要素及びその組み合わせの全てが本発明の解決手段に必須であるとは限らない。

[0016] (定義)

本明細書において、「解析対象RNA」又は「目的RNA」とは、互換的な意味を有し、生体内で低分子化合物やタンパク質と相互作用する可能性がある配列を有するRNA分子をいう。この解析対象RNAは、生体から抽出して得られた生物学的試料をそのまま用いてもよく、あるいは人工的に合成したRNAであってもよい。人工的に合成する場合は、RNAの配列情報に基づいて抽出した、RNAの機能構造単位であるモチーフ領域を含むことが好ましい。「モチーフ領域」とは、RNAが対象となる物質と相互作用するための機能構造単位を意味する。このRNAモチーフの構成要素であるステムループやシュードノットなどを構造モチーフと称し、この構造モチーフの組み合わせによってRNAの高次構造が形成される。本発明のRNAプローブに含まれるモチーフ領域は、単一のステムループ構造（ヘアピンループ構造）からなる場合もあれば、複数のステムループ構造（多分岐ループ構造）を含む場合もある。また1つ以上のキンクターン（kink-turn）、シュードノット（pseudoknot）、グアニン4重鎖（G-quadruplex）を含む場合もある。

u a d r u p l e x) などを含む場合もある。また構造モチーフはワトソン
クリック塩基対だけでなくフーグスティーン塩基対によっても構成され得る
。

[0017] 「RNAプローブ」とは、解析対象RNAを含む核酸分子、好ましくは、
RNAからなる核酸分子であって、増幅のためのプライマー結合部位やバー
コード配列などが付加された核酸分子をいう。また、「ライブラリ」とは、
複数（2つ以上）種類の異なる分子（例えば、複数の異なるDNA分子又は
複数の異なるRNA分子など）の集合を指す。本実施形態に係る方法では、
必要に応じて、多数のRNAプローブを用いて解析を行うことができるため
、用語、「ライブラリ」は、好ましくは10個以上、より好ましくは 10^2 個
以上、 10^3 個以上、又は 10^4 個以上、さらに好ましくは 10^6 個以上の異なる
RNA分子を含み得る。

[0018] 「RNAの高次構造」とは、溶液中において、主に、分子内での塩基対の
形成に基づく部分的な二本鎖形成（ステム構造ともいう）と、該塩基対形成
のない部分の一本鎖構造、又は環状一本鎖構造（ループ構造という）と、ま
たはそれらの組み合わせをいうものとする。このような構造は、溶液の状態
（温度、塩濃度等）により特定の平衡状態にありRNA分子の運動とともに
変動するものである。「ステム構造」とは、RNAに含まれる任意の核酸配
列と当該核酸配列に対して相補的な配列とにより形成される二重らせん構造
を意味する。本明細書において、「相補的」とは、2つの核酸配列がハイブ
リダイズする能力を意味し、2つの配列がハイブリダイズすればよいことか
ら、ステム構造を構成する2つの核酸配列は、少なくとも50%、55%、
60%、65%、70%、75%、80%、85%、90%、95%、99
%、または100%の配列相補性を有していればよい。

[0019] 「バーコード配列」とは、核酸分子に対して1種類毎に又は1分子毎に付
加される固有の配列を有するタグである。「インデックス」又は「固有分子
識別子（Unique Molecular Identifier: UMI）」などとも呼ばれる。UMIは、典型的には溶液中の各分子にランダム

配列を割り当てることで増幅バイアスの軽減による定量性の向上を目的としている。複数の解析対象RNAに対して、1種類のRNA毎に異なる固有の配列を有するバーコード配列が付加されるようにすると、複数のRNAを同時に修飾処理及び増幅処理した後に、付加されたバーコードの種類に基づいて、夫々のRNAを識別して解析することができることとなる。また、反応溶液、反応条件毎に異なるバーコードに基づいて各実験系列を識別することで、同一の次世代シーケンスデータから複数の実験データを分別して得られ、効率的なデータ解析が可能となる。

[0020] バーコード配列は、例えば、ランダムな塩基を有する核酸群として提供しうる。バーコード配列は、その配列の種類の数が必要なため、配列がランダム（配列が多様であり、かつ、配列の内容を認識する必要がないように）に合成されたものであったとしてもよい。あるいは、バーコード配列は、十分な多様性が得られるように設計された配列既知の核酸群であってもよい。

[0021] （RNAの高次構造を解析する方法）

図1は、本発明の一実施形態における、RNAの高次構造の解析方法を示すフロー図である。この方法は、解析対象RNAにバーコード配列を付加した1又は複数のRNAプローブを調製する工程（S10）と、RNAプローブとRNA修飾剤とを接触させる工程（S20）と、工程S20で得られたRNAプローブの配列中で修飾を受けた塩基の位置と頻度を検出する工程（S30）と、さらに必要に応じて、検出結果を表示する工程（S40）と、を含む。ここで、バーコード配列は、RNA修飾剤との反応が抑制される構造を有することを特徴とする。

[0022] また、別の実施形態におけるRNAの高次構造を解析する方法は、図2に示すように、上記検出工程（S30）が、工程S20で得られたRNAプローブの混合物を鋳型として逆転写酵素により相補DNAを合成する工程（S31）と、相補DNAの塩基配列を決定し、バーコード配列を含む塩基配列を整列させる工程（S32）と、整列させた塩基配列に生じた変異の位置と頻度を検出する工程（S33）と、からなることが好ましい。以下、これら

の各工程について詳細に説明する。

[0023] <RNAプローブの調製工程 (S10)>

解析対象RNAは、生体内での機能を発揮するためのモチーフ領域を含むことが好ましい。このモチーフ領域は、単一のステムループ構造（ヘアピンループ構造）からなる場合もあれば、複数のステムループ構造（多分岐ループ構造）を含む場合もある。本実施形態では、ステム構造を基準としてモチーフ領域を抽出することが好ましい（例えば、WO2018/003809明細書参照）。これにより、モチーフ領域を分断することなく、RNA中に実在する機能構造単位を反映したRNAプローブを調製することができる。モチーフ領域は、その機能が維持されていることを限度として、任意の配列長であってよく、例えば1000塩基以下、900塩基以下、800塩基以下、700塩基以下、600塩基以下、500塩基以下、400塩基以下、300塩基以下、200塩基以下、150塩基以下、100塩基以下、50塩基以下であってよい。

[0024] 上記解析対象RNAを含むRNAプローブの調製方法において、RNA中のステム構造は、例えばCentroidFold (Hamada, M. et al., Bioinformatics, Vol. 25, pp465-473, 2009) やIPknot (Sato, K. et al., Methods Biochem. Anal., Vol. 27, pp. i85-i93, 2011) などのRNA二次構造予測ソフトを用いて認識することができる。また、RNAの配列情報には任意のものを使用することができ、例えば、UTRdb (Grillo, G. et al., Nucl. Acids Res., Vol. 38, D75-D80, 2010)、IRESite (Mokrejs, M. et al., Nucl. Acids Res., Vol. 38, D131-D136, 2010)、GenBank (Benson, D. et al., Nucl. Acids Res., Vol. 41, D36-D42, 2013)、RNAcentral (RNAcentral Consortium, Nucl. Acids Res.

s., Vol. 43, D123-D129, 2015)などのRNA配列データベースからダウンロードしたものを使用することができる。また、RNAの配列情報だけでなく構造情報も含むデータベースからRNAの配列情報を入手してもよく、例えば、Rfam (Nawrocki, E. P. et al., Nucl. Acids Res., Vol. 43, D130-D137, 2015)、Structure Surfer (Berkowitz, N. D. et al., BMC Bioinformatics, Vol. 17, p. 215, 2016)などからダウンロードしたものを使用することができる。また種々の手法によって決定されたRNAの三次元構造データを使用してもよく、例えば、Protein Data bank (<https://www.rcsb.org/>)などからダウンロードしたものを使用することができる。また自身で設計をしたRNA高次構造でもよく、例えば、RNAinverseなどのソフトウェアで設計されたデータを使用しても良い。

[0025] 本工程では、解析対象RNAに付加されるバーコード配列が構造化されていることを特徴とする。「構造化」とは、RNAが溶液中で折りたたまれて二次、三次構造を形成するか、あるいは一次構造（配列）の状態に留めることをいい、それによりRNA修飾剤との反応が抑制される。例えば、RNA修飾剤が、RNAプローブ中の一本鎖領域のような非拘束ヌクレオチドを選択的に修飾する化合物であるとき、バーコード配列は、このような修飾を受けにくい複数の塩基対を含む構造を形成するように設計することができる。複数の塩基対とは、連続又は離間する2以上の塩基がバーコード配列内の他の塩基との間で水素結合を形成することを意味し、ワトソンクリック型塩基対又はフーグスティーン型塩基対等のいずれでもよい。ワトソンクリック型塩基対と同程度の熱力学的安定性を有するG-Uゆらぎ塩基対でもよい。生体内と同様の環境又は条件において、2以上の塩基対があれば少なくとも一時的には安定な構造を形成しうるが、より安定な構造を形成するためには3個以上の塩基対が好ましく、4個以上の塩基対がさらに好ましく、5個

以上の塩基対がさらになお好ましい。塩基対の個数の上限は特に制限されないが、10個程度の塩基対があれば十分に安定な構造が得られるため、コスト的な観点から30個以下の塩基対が好ましく、20個以下の塩基対がより好ましく、15個以下の塩基対がさらに好ましい。一方、RNA修飾剤が、RNAプローブ中で二本鎖を形成するような拘束ヌクレオチドを選択的に修飾するときは、バーコード配列は、塩基対を形成しない配列、すなわち、一本鎖の構造を維持するように設計されていることが好ましい。さらに、この構造を有するバーコード配列は解析対象のRNAに影響を与えないように計算機的に配列を最適化することが好ましい。これはバーコード配列の付与自体が解析対象のRNAと分子内相互作用により元来のRNA構造とはかけ離れた構造を形成するか、あるいは構造の安定性に影響がある問題を回避するためである。計算機的な配列最適化は、Vienna RNAパッケージ等の公知のプログラムを用いて行うことができる。

[0026] <RNAプローブの修飾工程(S20)>

本工程(S20)におけるRNAの修飾反応は、前工程(S10)で調製したRNAプローブと、所望のRNA修飾剤とを接触させることでRNAプローブの修飾反応を起こさせるものである。1つの実施形態として、このRNA修飾剤は、RNAプローブ中の一本鎖領域のような非拘束ヌクレオチドを選択的に修飾する化合物が挙げられる。このような化合物は、典型的には、SHAPE試薬として知られる、リボース-2'-ヒドロキシ基と反応するイサト酸無水物誘導体、例えば、1-メチル-7-ニトロイサト酸無水物(1M7)、1-メチル-6-ニトロイサト酸無水物(1M6)、NMIA(N-メチルイサト酸無水物)及び2-メチルニコチン酸イミダゾリド(NAI)を含むがこれらに限定されない。SHAPE試薬の他に、硫酸ジメチル(DMS)は、アデノシンのN1位置、シトシンのN3位置、及びウリジンのN3位置、グアノシンのN1位置で付加物を形成するため、RNA修飾剤として用いることができる。一例として、NAIは一般的に4つ全てのヌクレオチドと反応し、DMSは、アデニンとシトシンのみと反応する。一方

で、DMSは塩基性に偏ったpH（例えばpH8.0）条件下にてグアニンとウリジンにも反応できる。

[0027] 他の実施形態として、RNA修飾剤は、RNAプローブ中で二本鎖を形成するような拘束ヌクレオチドを選択的に修飾するものであってもよい。このRNA修飾剤は、例えば、二本鎖RNAを分解する酵素であるRNASE V1や、RNASE IIIファミリーのDICER、または二本鎖結合タンパク質とRNA修飾タンパク質との融合タンパク質などが含まれるがこれらに限定されない。

[0028] このようなRNA修飾剤を含む溶媒溶液を、RNAプローブを含む溶液に添加して、RNAプローブとRNA修飾剤とを接触させる。この溶液は、異なる濃度及び量のタンパク質、細胞、ウイルス、脂質、単糖及び多糖類、アミノ酸、ヌクレオチド、DNA、並びに種々の塩及び代謝産物を含有する生体溶液であってもよい。また異なる濃度及び量の低分子、中分子薬剤を含有する溶液であっても良い。また種々の界面活性剤、ポリマー、オズモライトを含有しても良い。RNA修飾剤の濃度は、RNAに所望の修飾の程度を達成するように調整することができる。

[0029] さらに、解析対象RNAは、タンパク質又は他の低分子及び高分子の生物学的リガンドの存在下で修飾することができる。RNA修飾剤の反応性がpHに依存する場合には、そのpHを、例えば、7.5～9.0の範囲、但しこれに限定されない、に維持してもよい。最大の反応性と最小の反応性ヌクレオチドを区別する機能範囲は、典型的には20～50倍に及ぶ。RNAを所望のpH（例えば、約pH8）で所望の立体配座に折り畳む任意の手順で置換することができる。このRNAを、多量体形態を排除するために、まず加熱し、続いて急いで低イオン強度緩衝液中で冷却することができる。続いて、フォールディング溶液を加えて、RNAが正確な立体配座を達成し、構造に感度が良いRNA修飾剤で探るための準備とすることができる。いくつかの実施態様において、RNAは、修飾前には自然には折り畳まれていない。RNAが熱及び／又は低塩条件により変性されている間に、修飾が行われ

得る。

[0030] <修飾塩基の検出工程 (S 3 0) >

本工程は、上記修飾工程 (S 2 0) で得られたRNAプローブの配列中で、修飾を受けた塩基の位置と頻度を検出する工程である。RNA配列中における修飾塩基を読み取る方法であれば特に限定されず、例えば、修飾塩基に特異的な抗体を用いるプルダウン法や直接RNAの電位を読み取るナノポアシーケンス法であってもよい。この直接RNAナノポアシーケンス法は、単一分子レベルでRNAの修飾部位を検出するための技術である。現在、Oxford Nanopore Technologiesが開発及び市販している直接RNAシーケンシングプラットフォームでは、膜に懸濁された生物学的ナノポアを介してモータータンパク質と結合したRNAが移動する。RNAが電圧バイアス下で細孔を通過するとき、細孔狭窄部を通過する短い配列 (5ヌクレオチド) の化学的同一性 (つまりシーケンス) に依存して、ピコアンペアのイオン電流の変化が観察される (Garalde, D. R., et al. (2018) Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods, 及びWorkman, R. E., et al. (2019) Nanopore native RNA sequencing of a human poly (A) transcriptome. Nat. Methods, 16, 1297-1305. 参照)。SHAPE試薬の1つである、1-アセチルイミダゾール (AcIm) により修飾されたヌクレオチドを、この方法で検出しうることが報告されている (William Stephenson et al., Direct detection of RNA modifications and structure using single molecule nanopore sequencing. bioRxiv doi:https://doi.org/10.1101/2020.05.31.126763, Posted June 01, 2020)

。

[0031] 好ましい実施形態において、修飾塩基の検出工程（S30）は、図2に示すように、RNAから相補DNA（cDNA）への変換を含む変異プロファイリングである。この実施形態では、最初に、工程S20で得られたRNAプローブの混合物を鋳型として、逆転写酵素又は他のポリメラーゼによりcDNAを合成する（S31）。逆転写酵素とは、RNAからcDNAを合成する酵素であり、例えば、マウス又は鳥類の逆転写酵素のような熱安定性酵素が挙げられるが、これらに限定されない。あるいは、原核生物や真菌などのレトロトランスポゾン内に存在する逆転写酵素TGIRT（Thermally Stable Group II intron reverse transcriptase）であってもよい。InGex社のTGIRT-IIは、従来のレトロウィルス由来の逆転写酵素に比べて、熱安定性や処理性、正確性が優れている。またDMSによる修飾箇所において逆転写時に変異を誘発する性質が知られている（DMS-MaPseq法）。

[0032] これらの酵素は、付加物を含むヌクレオチドを読み飛ばし、化学修飾の部位に不正確な（非相補的な）ヌクレオチドを組み込ませることにより、RNA中の化学的修飾を検出する方法が含まれる。本明細書で使用される、ヌクレオチドの取り込みに関する、「不正確な」とは、元の配列中に存在するヌクレオチドに、非相補的なヌクレオチド（ワトソン-クリックの規則に反するヌクレオチド）を組み込むことをいう。これは配列内の少数の欠失を含む。

。

[0033] 続いて、cDNAの塩基配列を決定し、バーコード配列を含む塩基配列を整列させる（S32）。cDNAは、多種類のRNAプローブの混合物に由来するライブラリを用いることにより、超並列配列決定法（MPS）を使用して、RNAなどの核酸中の化学修飾を効率よく検出することができる。一例として、イルミナ社の次世代シーケンサーでは、数千万～数億のDNA断片の両端部のアダプターを介して5'末端側をフローセル上に固定させる。次に、フローセル上に予め固定された5'末端側のアダプターと、DNA断

片の3'末端側のアダプター配列とをアニールさせて、ブリッジ状のDNA断片を形成させる。この状態でDNAポリメラーゼによる核酸増幅反応を行うことで、多数の1本鎖DNA断片を局所的に増幅して固定することができる。そして、次世代シーケンサーでは、得られた1本鎖DNAを鋳型として、シーケンシングを行うことで、2020年現在、1回の解析において約3 Tbという膨大な配列情報を得ることができる。核酸を高速かつ並列に判読するこれらの技術は「次世代シーケンシング (Next-Generation Sequencing (NGS))」、「超並列シーケンシング」、「超高処理量遺伝子シーケンシング (Ultra-High-Throughput Sequencing)」又は「大規模並列シーケンシング」などとも呼ばれる。

[0034] 1つの実施形態では、次世代シーケンサーによって得られた配列データ（リード）を、バーコード配列を含んだ形で整列させる。個々のバーコード配列ごとに配列データを整列させることにより、多種類のRNAプローブを含むサンプルを同時にシーケンスすることができるからである。また、解析対象RNAが類似する配列、例えば、遺伝子ファミリー、一塩基多型などを含む場合であっても、それらを識別して解析することが可能となる。

[0035] あるいは、すべてのcDNAと一緒に整列させた後、信頼度の低い整列物に関して、バーコードの変異情報を加味することでその整列を評価しても良い。いずれの方法においても、バーコード配列とともに解析対象RNA配列を整列させることで配列情報の正確性を向上することができる。

[0036] このように整列させた塩基配列に基づいて、生じた変異の位置と頻度を検出する（S33）。所定のヌクレオチドにおける変異率は、単純にその場所において、変異数（ミスマッチ、欠失及び挿入）を読み取り回数で割ったものである。各ヌクレオチドについて生の反応性を算出したデータは、種々の基準を用いて正規化することができる。シーケンスの読み取り深度や標準誤差を考慮することでデータの品質管理が可能である。

[0037] <検出結果の表示（S40）>

上記工程で検出された変異の位置と頻度は、変異ヒストグラム、シーケンスの深さ及び反応性プロファイルなどの当業者に既知の方法で図示することができる。変異位置と頻度の解析はBWA、STARなどの整列用ソフトウェア（アラインメントソフトウェア）を使用することができる。それらのデータは変異カウントとして数値化、ベクトル化され種々の演算を実施できる。また、統計的優位な反応性を示した変異に対してアノテーションをつけることができる。

[0038] 本工程におけるこれらの解析は、コンピュータ読取り可能媒体に保存されたコンピュータプログラム製品を用いて実施することができる。本発明を実施するために適当な例示的コンピュータ読取り可能媒体には、チップメモリデバイス、ディスク記憶装置、プログラマブルロジックデバイス、及び特定用途向け集積回路が含まれる。更に、本工程を実施するコンピュータプログラム製品は、単一のデバイス又はコンピューティングプラットフォーム上に設置することができ、又は複数のデバイス又はコンピューティングプラットフォーム間に分散させることができる。従って、本実施形態の方法により取得したRNAの高次構造をコンピュータと接続されたディスプレイ上に表示することができる。

[0039] （作用効果）

本実施形態で開示した構造化バーコードは、いくつかの有利な作用効果を有する。1つは、RNA修飾剤との反応において、バーコード配列が修飾される可能性が低く、バーコードとして正しく識別することが可能となる。またバーコード部分が解析対象RNA又は他のRNA分子と相互作用することが抑制される。これにより、構造化バーコード配列は、ライブラリ内の類似配列と識別できるだけでなく、同じライブラリの異なるバッチを区別することも可能である。例えば、図4は、37種類の第1のバーコード配列と、4種類の第2のバーコード配列を用いてライブラリ群を作製する方法を表す。最初に作製された37種類のDNAからなるライブラリを4種類の異なるプライマーを用いて増幅することで、1つのライブラリ内では同じ配列である

が、異なるバッチのライブラリでは異なる配列を有する第2のバーコード配列が付加される。これらを用いてインビトロ転写反応を行うことで、2種類のバーコード配列が付加されたRNAライブラリ群を作製することができる。

[0040] (RNAプローブ及びRNAプローブライブラリ)

本発明の他の実施形態としては、構造化されたバーコード配列を含むRNAプローブ及び複数の当該RNAプローブを含むRNAプローブライブラリが提供される。1つの実施形態において、構造化バーコード配列とは、複数の塩基対を含む構造を形成するバーコード配列である。本実施形態のバーコード配列としては、例えば、相補的な二本鎖構造、三重鎖構造又は四重鎖構造を含み、具体的には、ステムループ構造、シュードノット構造などを挙げることができる。ステム部分は相補的な二本鎖を形成するが、配列の多様性を増やすために、ワトソンクリック型塩基対と同程度の熱力学的安定性を有するG-U、I-U、I-A及びI-Cのゆらぎ塩基対 (wobble base pair) を含んでいてもよい。Iは、イノシンを表し、その塩基であるヒポキサンチンはウラシル、アデニン、シトシンと塩基対形成が可能である。ウラシルはグアニンとアデニンという2種類の塩基と対合することが可能である。

[0041] 他の実施形態では、複数の塩基対を含む構造が、ステムループ構造であり、ステム部位に1つ以上のバルジ及び／又は内部ループ構造を有する。これにより構造化バーコードがとり得る高次構造の多様性、種類を向上させることができる。また、構造特異的な変異のネガティブコントロール、ポジティブコントロールとなる塩基を同時に搭載することができる。また1本鎖RNA修飾剤のうち、末端ループには修飾を行うが、バルジや内部ループには修飾をしない分子種に対して構造化バーコードがコントロールとして機能する。

[0042] いくつかの実施形態としては、複数の塩基対を含む構造が、PDB (Protein Data Bank) に登録されているRNA構造又はその改

変体である。これによりワトソンクリック型ではないRNA高次構造をバーコード化することができる。例えば、これにより非ワトソンクリック型塩基対でないRNA高次構造に対する修飾剤に対する構造化バーコードとして有用である。

[0043] 本実施形態のRNAプローブ中における構造化バーコード配列の位置は特に制限されず任意の位置に配置することができる。例えば、解析対象RNAの5'末端側であっても3'末端側であってもよい。あるいは、相補鎖を形成するバーコード配列の一方の鎖が解析対象RNAの5'末端側に位置し、他方の鎖が3'末端側に位置してこれらが解析対象RNAを挟むように二本鎖を形成してもよい。また、構造化バーコード配列の数も特に限定されず、同一又は異なる配列の構造化バーコードが複数個存在してもよい。

[0044] 本実施形態のRNAプローブには、解析対象RNAとして少なくとも1つの構造モチーフを含むRNAモチーフが含まれる。このモチーフ領域には、任意のRNA配列情報から抽出したものをを用いることができる。または、本発明のRNAプローブに含まれるモチーフ領域には、RNAストラクチュローム研究によりすでに特定された任意のRNA二次構造データから選択されたものをを用いてもよい。

[0045] さらに、このRNAプローブは、検出のために、蛍光色素（例えば、FITC、PE、Cy3、Cy5など）、放射性同位体、ジゴキシゲニン（DIG）、ビオチンなどにより標識されてよい。標識は、予め標識した核酸をプローブ合成時に取り込ませることによって行うことができ、例えば、5'側に標識された人工核酸を取り込ませることができる。またRNA全長に標識された人工核酸を取り込ませることができる。3'側には例えばT4 RNA ligase 1を用いて標識された人工核酸を標識することができる。標識は、クリック反応などによって多段階的に行われてもよい。例えば、pCp-N3をT4 RNA ligase 1を用いて3'末端に付与したRNAに対して、DBCO-biotin、DBCO-Cy3を反応させることでRNAに蛍光色素やビオチンを取り込ませることができる。これら標識

の割合は10, 20, 30, 40, 50, 60, 70, 80, 90, 99, 100%でも良い。

[0046] 本実施形態のRNAプローブは、従来公知の任意の遺伝子工学的方法により合成することができる。好ましくは、RNAプローブは、合成の受託業者に委託して合成された鋳型DNAを転写することによって作製することができる。DNAからのRNAへの転写を行うため、RNAプローブの配列を含むDNAは、プロモーター配列を有していても良い。特に限定されないが、好ましいプロモーター配列として、T7プロモーター配列が例示される。T7プロモーター配列を用いた場合、例えば、ライフテクノロジーズ社より提供されるMEGAshortscript (商標) T7 Transcription Kitを用いて所望のRNAプローブ配列を有するDNAよりRNAを転写行うことができる。本発明において、RNAは、アデニン、グアニン、シトシン、ウラシルのみならず、修飾RNAであってもよい。修飾RNAは、例えば、プソイドウリジン、5-メチルシトシン、5-メチルウリジン、2'-O-メチルウリジン、2-チオウリジン、およびN6-メチルアデノシンが例示される。

[0047] 1つの実施形態において、それぞれ異なる配列の解析対象RNAを含む複数のRNAプローブを含むRNAプローブライブラリを提供する。本実施形態では、多種類のRNAプローブを同時に用意することが好ましく、効率的にRNAプローブの鋳型を含有するオリゴ核酸ライブラリ合成 (Oligonucleotide Library Synthesis) 技術を用いて行うことが好ましい。これは、スライド上の規定した位置に個々の塩基をプリントするインクジェット技術を用いて、一度に1塩基ずつ合成して、指定の長さの鋳型DNAを伸長させる。次に、構築されたオリゴはスライドから切断され、プール化された後、乾燥し、1本のチューブに入れて保存される。オリゴライブラリは、その後、再溶解、増幅後、インビトロ転写反応によりRNAプローブライブラリを調製することができる。本発明において特に限定されないが、Oligonucleotide Library S

y n t h e s i s は、アジレントテクノロジー社やツイスト・バイオサイエンス社に委託することによって作製することができる。

[0048] さらに他の実施形態では、複数のRNAプローブを含む本実施形態のRNAプローブライブラリを、第2のバーコード配列を含む複数のプライマーを用いて増幅することにより、2以上の複製物からなるRNAプローブライブラリ群を調製することができる。複製されたすべてのRNAプローブは、第1及び第2のバーコード配列を含み、この第2のバーコード配列は、1つのライブラリ内ではすべて同一配列であるが他のライブラリとの間では識別可能である。後述する実施例によれば、複数のRNAプローブライブラリを混合して変異プロファイリングを行った場合でもそれぞれのRNAプローブライブラリを用いて行った場合と同じ結果が得られたことから、各RNAプローブライブラリを用いてそれぞれ異なる変異プロファイリングを行った後、これらを混合して次世代シーケンシングを行った後、第2のバーコード配列を用いてそれぞれの変異プロファイリングを識別することができると考えられる。したがって、第2のバーコード配列を付与することで、反応条件の異なる変異プロファイリングを行った場合でも同一条件で次世代シーケンシング用ライブラリを調製することができ、変異のプロファイリング工程の格段の効率化が可能となる。この第1及び第2のバーコード配列を含むRNAプローブライブラリ群に、さらに異なるバーコード配列を付加することで3種類又はそれ以上のバーコード配列を含むRNAプローブライブラリ群を作製することも可能である。

[0049] 本実施形態のRNAライブラリは、RNAの化学修飾の分析及び／又はRNA構造分析を行うためのキットとして使用することができる。このようなキットの使用方法として、本発明に係るRNAの高次構造解析方法が含まれる。次に実施例を挙げ、本発明を更に詳しく説明するが、本発明はこれら実施例に何ら制約されるものではない。

実施例

[0050] [実施例1]

材料と方法

(バーコード配列の設計)

本実施例におけるバーコード配列は、異なる長さのステムとループを使用した。正規の塩基対とGUゆらぎ塩基対を含む、長さ6、7又は8塩基対 (bp) のステムをランダムに生成した。ステムの長さごとに、3つの異なる長さのループを使用した。各バーコードに対して、4つのテトラループ (UUCG、GAGA、GCUU、GUAA) のいずれか1つ、又は3もしくは5塩基長の配列 (UUCG、AGA、CUU、UAA、UUACG、GAAGA、GCUAU、AGUAA) のいずれか1つを選択した。ViennaRNAパッケージを使用して、バーコードを正しく折りたたむように制御した。コントロールとして、構造化されていない10、15及び21塩基長のバーコードを生成した。

[0051] (目的RNA配列)

構造化バーコードの有用性を実証するために、目的RNAとして以下の配列：

5' -GUGUAUGAUGAAACUACA UUAAGUUAACUCG
UGCAC-3' (配列番号1) を用いた。この配列から、塩基対を形成しない12カ所の位置を選択し、各位置において、他の3つすべての塩基に変えた点変異体を作成することにより、36個の点変異体を得た。これにより、合計37個の配列が得られた。この37個の配列の任意のペアは、1又は2塩基のみが相違する。

[0052] (第1のライブラリ設計)

第1のライブラリに用いたバーコード配列及びライブラリ構造の概要を図3に示す。図3(a)は、1つのRNAプローブ(ID1)のバーコード配列であり、7bpのステムと4ヌクレオチドのループで構成されている。第1のライブラリ配列は、5' から3' の方向に以下の4つの部分を有する：

i) インビトロ転写(IVT)によるRNAライブラリの生成と、シーケンス用ライブラリの調製に必要な5' カセット (図3(b)における5' 側の

破線) ;

i i) 個々の配列ごとに異なるバーコード配列 (図3 (b) の構造化バーコードを含むID 1 ~ 28 及び非構造化バーコードを含むID 29 ~ 37) ;

i i i) 両側に2塩基のスペーサーが隣接する目的RNA配列 (図3 (b) の実線、なお、配列中の点変異を三角形で示す。) ;

i v) インビトロ転写 (IVT) によるRNAライブラリの生成、逆転写及びシーケンス用ライブラリの調製に必要な3' カセット (図3 (b) における3' 側の破線) 。

[0053] (第2のライブラリ設計)

第2のライブラリに用いたバーコード配列及びライブラリ構造の概要を図4に示す。この設計によるRNAは、ライブラリ内バーコード(第1のバーコード)とバッチバーコード(第2のバーコード)の2つのバーコードを含む。5' から3' の方向に以下の4つの部分に分けることができる:

i) 第1のライブラリ設計で用いたものと同じ5' カセット;

i i) 第1のライブラリ設計で用いたものと同じバーコード配列;

i i i) 両側に2塩基のスペーサーが隣接する目的RNA配列;

i v) プライマー結合を強化する12塩基のリンカー配列。

[0054] インビトロ転写 (IVT) を行う前に、さらに以下の2つの部分をポリメラーゼ連鎖反応 (PCR) によって取り付けた。

v) 4種類のバッチバーコード。このバーコードは、1つのバッチ内のすべての目的RNAで同じ配列である。

v i) 第1のライブラリ設計で用いたものと同じ3' カセット。

[0055] 第2のライブラリの増幅のために用いたプライマーの塩基配列は以下のとおりである。

[表1]

名称	配列	配列番号
Pr_d2a	5'-GAACCGGACCGAAGCCCGATTTGGAGCGGAAGCCGCTCACCGTT GACCAGTTGTGCAC-3'	2
Pr_d2b	5'-GAACCGGACCGAAGCCCGATTTGGGCCACAAGGTGGCCACCGTT GACCAGTTGTGCAC-3'	3
Pr_d2c	5'-GAACCGGACCGAAGCCCGATTTGACCTCTGTTACAGAGGTACCG TTGACCAGTTGTGCAC-3'	4
Pr_d2d	5'-GAACCGGACCGAAGCCCGATTTGACACACCTTAGGTGTGTACCG TTGACCAGTTGTGCAC-3'	5

[0056] (DNA鎖の合成)

上述したライブラリ及びプライマーは、DNAの形でIntegrated DNA Technologies, Inc. (IDT社)に依頼して合成した。コントロールとして、第1のライブラリで設計した構造化又は非構造化バーコード配列を持つ2つの個別のRNAプローブ(それぞれID1及びID32)を合成した。

[0057] (DNAからRNAの合成)

まず、Platinum™ SuperFi™ PCR Master Mix (サーモフィッシャーサイエンティフィック株式会社製)を使用して、ライブラリをPCRで増幅した。第1のライブラリと、このライブラリ中の2つの個別の一本鎖RNA用には、T7 RNAポリメラーゼプロモーター配列(IVTの認識サイト: 5'-TAATACGACTCACTATAG-3' (配列番号6))の下流に5'カセット配列を有するフォワードプライマーと、3'カセット配列に相補的な配列を有するリバースプライマーを使用した。第2のライブラリを調製するためのリバースプライマーとしては、Pr_d2a (配列番号2)、Pr_d2b (配列番号3)、Pr_d2c (配列番号4)及びPr_d2d (配列番号5)を使用して4つの異なるバッチを作成し、バーコードを付加した。すべての反応において、各プライマーは、最終濃度500 nMになるように添加し、テンプレートは総濃度0.4 nMで提供した。反応容量は25 µLであった。すべてのPCRはサーモフィッシャーサイエンティフィック株式会社のProFlex™ PCRシステ

ムで行った。

[0058] 30秒間98℃に最初に加熱した後、98℃で10秒間、68℃で10秒間、72℃で15秒間の3ステップPCRを行った。最後のサイクルの後、温度は72℃で5分間保持され、その後4℃に冷却した。2.5μLのエキソヌクレアーゼI (New England Biolabs Inc.) を各チューブに加え、37℃で15分間インキュベートした後、再び4℃に冷却した。精製には、Monarch (登録商標) PCR&DNA Cleanup Kit (5μg) (New England Biolabs Inc.) のDNAクリーンアップおよび濃縮プロトコルを使用した。最終溶出には、10μLのDNA溶出バッファーを使用した。サーモフィッシャーサイエンティフィック株式会社のNanoDrop™Oneを使用して、最終濃度を測定した。

[0059] 調製された二本鎖DNAは、MEGAshortscrip™T7転写キット (サーモフィッシャーサイエンティフィック株式会社) を使用したIVT反応のテンプレートとして使用した。反応はマニュアルに従って調製した。反応容量は20μL、テンプレート濃度は100nMである。反応物を37℃で6時間インキュベートした後、TURBO DNase (キットに付属) で、37℃で15分間処理した。次に、RNAをZymo ResearchのRNA Clean&Concentrator-25で精製した。

[0060] インビトロ転写反応により合成した第1のライブラリに含まれるRNAプローブのうち、個別のストランドとして合成したID1 (配列番号7) 及びID32 (配列番号8) の塩基配列を図5に示す。図5において、それぞれのバーコード配列部分を四角で囲み、目的RNA配列には下線を付した。

[0061] (構造プロファイリング用のRNAの調製)

RNA修飾には2つの異なる化学修飾剤を使用した。シグマアルドリッチから購入したメチル化剤の硫酸ジメチル (DMS)、及びSHAPE試薬2-メチルニコチン酸イミダゾリド (NAI) である。両方の修飾剤を用いた

実験では、同じRNA調製物を使用した。6 μ Lの水に溶解した250 ngのRNA（一本鎖またはプール）を95℃で2分間インキュベートし、氷上で少なくとも2分間急冷した。次に、3 μ Lの3. 3×フォールディングバッファを加え、サンプルを37℃で20分間インキュベートした（1×フォールディングバッファは、100 mM HEPES（pH 8. 0），100 mM NaCl，10 mM MgCl₂で構成されている）。

[0062] （NAIによる構造プロファイリング修飾）

1000 mMのNAI溶液1 μ Lを、空の0. 2 mLのPCRチューブに加えた。RNAを加える直前まで、チューブを氷上で維持した。37℃で、RNAを含む9 μ LのサンプルをNAIに加え、溶液を上下にピペッティングして混合した。サンプルは37℃で10分間放置した。

[0063] 反応停止後、RNAをZymo Research社のRNA Clean and Concentrator-5キットで精製し、最終溶出量を15 μ Lにした。NAIで修飾された各RNAサンプルについて、NAIの代わりに1 μ LのDMSOを使用して同じ方法で処理したコントロールサンプルを調製した。

[0064] （DMSによる構造プロファイリング修飾）

37℃で、エタノールを含む1 μ Lの50%DMSを、先に調製したRNAを含む9 μ Lのサンプルに加えた。サンプルを37℃で6分間放置した。5 μ Lの β -メルカプトエタノールで反応を停止し、完全に混合した後、37℃で2分間インキュベートした。次に、RNAをZymo ResearchのRNA Clean and Concentrator-5キットで精製し、最終溶出量を15 μ Lにした。DMSで修飾された各RNAサンプルについて、DMSの代わりに1 μ Lの50%エタノール水溶液を用いて同じ方法で処理したコントロールサンプルを調製した。

[0065] （変異プロファイリングのための逆転写）

修飾されたRNAサンプルは、3' カセット配列に相補的な配列を有するリバースプライマーを使用して逆転写反応を行った。NAI修飾RNAの場合

合、マンガンの存在下で酵素SuperScript™ II 逆転写酵素（サーモフィッシャーサイエンティフィック株式会社）を使用した。DMS修飾RNAの場合、TGIRT™-III 酵素（InGex）を使用した。どちらの場合も、1 μ Lの2 μ Mリバースプライマーを2 μ Lの10 mM dNTP（New England Biolabs）と7 μ Lの先に修飾したRNAと混合した。サンプルは、サーモフィッシャーサイエンティフィック株式会社のProFlex™ PCRシステムでアニールされ（85℃、1分→65℃、10分→4℃で保持）、これは逆転写ステップにも使用した。次に、9 μ Lの2.22×MaPバッファーを添加して、室温で2分間インキュベートし、1 μ Lの酵素を加え、サンプルをサイクラーに入れて逆転写した（表2を参照）。

[0066] [表2]

	DMS処理したRNA	NAI処理したRNA
1xMaP-buffer	100 mM Tris (pH 8.0), 75 mM KCl, 3 mM MgCl, 1M betaine, 10 mM DTT	50 mM Tris (pH 8.0), 75 mM KCl, 6 mM MnCl ₂ , 1M betaine, 10 mM DTT
酵素	TGIRT™-III	SuperScript™ II Reverse Transcriptase
温度プロトコル	25℃ 5min → 57℃ 90min → 90℃ 10min → 4℃ hold	25℃ 10min → 42℃ 90min → [50℃ 2min → 42℃ 2min] x 15 → 72℃ 10min → 4℃ hold

[0067] 逆転写が終了したら、1 μ LのRNase Hを加え、サンプルを37℃で20分間インキュベートした。精製には、AMPure XP（Beckman Coulter製）、使用してプロトコルに従って精製した。溶出のために、乾燥したビーズに14 μ Lの水を加え、十分に混合し、室温で10分間インキュベートし、12.5 μ Lの上澄みを回収した。

[0068] （次世代シーケンシング用ライブラリの調製）

ライブラリの準備には、アンプリコンPCRとインデックスPCRの2つのPCRを行った。アンプリコンPCR用1 ngの逆転写生成物は、25 μ Lの反応容量で使用した。その他の反応コンポーネントは、1×Platinum™ SuperFi™ PCR Master Mixと1×SuperFi GC Enhancer（どちらもサーモフィッシャーサイエンティフィック株式会社製）、500 nMのフォワードプライマー及びリバースプ

ライマーを用いた。サンプルをProFlex™PCRシステムに移した。最初に、30秒間98℃に加熱した後、98℃で10秒間、64℃で10秒間、72℃で20秒間の3ステップPCRを行った。最後のサイクルの後、温度は72℃で5分間保持され、その後4℃に冷却した。精製には、Monarch（登録商標）PCR&DNA Cleanup Kit（5 μg）（New England Biolabs Inc.）のDNAクリーンアップおよび濃縮プロトコルを使用した。最終溶出には、8 μLのDNA溶出バッファーを使用した。これで、次世代シーケンシング用のインデックスを付ける準備ができた。

[0069] 次に、25 μLの反応容量で1 ngのアンプリコンPCR産物を用いてインデックスPCRを行った。その他の反応コンポーネントは、1x Platinum™ SuperFi™ PCR Master MixとNextera XT Index Kit v2（Illumina）の1 μMインデックスプライマーである。サンプルをProFlex™PCRシステムに移した。30秒間98℃に最初に加熱した後、98℃で10秒間、55℃で10秒間、72℃で20秒間の3サイクルPCRを6サイクル行った。最後のサイクルの後、温度は72℃で5分間保持され、その後4℃に冷却した。精製には、AMPure XP（Beckman Coulter製）を使用してクリーンアップした。溶出のために、乾燥したビーズに13 μLの水を加え、十分に混合し、室温で10分間インキュベートし、12 μLの上澄みを回収した。その後、サンプルは次世代シーケンシングのために一緒に混合した。

[0070] （次世代シーケンシング）

シーケンシングには、ペアエンドリードと標準リードプライマーを使用したNextSeq 500/550ミッドアウトプットキットv2.5（イリミナ社、150サイクル）を使用した。

[0071] （アラインメントとデータ分析）

FASTQファイルのアダプターは最初にトリミングされ、次に、アライ

ンメントソフトウェアを使用して生成されたFASTQファイルのリードを、アラインメントソフトウェアを使用して参照配列が含まれたファイル（リファレンスファイル）に対してマッピングを行った。本解析ではSTARアライナーソフトウェアを用いてマッピングした。さらなる分析のために、変異、欠失および挿入をカウントした。

[0072] 図6は、第2のライブラリを用いて行った変異プロファイリング操作の流れを示す模式図である。それぞれ別々に化学修飾を行った4つのライブラリを1本のチューブにまとめて逆転写反応を行った。一方で比較対照用のサンプルとして前記4つのライブラリに対して別々に逆転写反応を行った4本のチューブを用意した。

[0073] 結果と考察

（RNAライブラリ内の配列を区別するためのバーコード）

バーコードが変異プロファイリング実験で類似の配列を区別するのに役立つかどうかをテストするために、第1の設計によるライブラリを使用した。文字列の類似度を測る指標としてレーベンシュタイン距離を使用して、2つの配列の類似性を測定した。この距離は、ある配列を別の配列に変換するための挿入、削除、変異の最小数を示す。バーコードを付加しなければ、ライブラリ内の配列の任意のペアに対して、この数は1又は2となる。バーコードを付加すると、レーベンシュタイン距離は7以上である。したがって、変異プロファイリング実験で予想される変異率の増加があっても、シーケンスを正しく識別することができる。完全なライブラリに加えて、ライブラリの2つの単一シーケンス（ID1とID32）をコントロールとして用いた。ID1は構造化バーコードを含むが、ID32は非構造化バーコードを含む（図5参照）。

[0074] 3つのRNAサンプル（ライブラリ及び2つコントロール）を、上述したようにNAI又はDMSで修飾した。NAI／DMSで処理しないコントロール実験を含め、合計12サンプルとなる。単一配列を含め上記サンプルを第1のライブラリのすべての配列に整列させた。データは、それぞれのID

について、各ヌクレオチドの変異 (M)、削除 (D) 及び挿入 (I) をカウントすることによって分析した。変異率 (mut) は、M、D、I を合計し、ある塩基位置での読み取りの総数で割ることによって計算した。配列固有の変異によるノイズを減らすために、未修飾サンプルの変異率を NAI / DMS 修飾サンプルの変異率から差し引いて、下記式 (1) のデルタ変異率を求めた。

$$\text{デルタ変異率} = \text{修飾変異率} - \text{未修飾変異率} \quad (1)$$

[0075] その結果を図 7 に示す。図 7 (a) は、NAI で修飾した第 1 のライブラリのバーコード配列中のすべてのヌクレオチドに関するデルタ変異率絶対値を示す箱ひげ図である。図 7 (b) は、DMS で処理したサンプルを同様に分析した結果である。図 7 において、ノッチは中央値を示し、ボックスは四分位範囲を示す。また、ひげは、ボックスの端からボックスの高さの 1.5 倍の範囲内にある最大値又は最小値まで上下する。外れ値は円で示した。これらの結果は、NAI 又は DMS のいずれの修飾剤で処理した場合でも、構造化バーコード配列 (ID 1 ~ 28) のデルタ変異率の絶対値が、非構造化バーコード配列 (ID 29 ~ 37) のデルタ変異率の絶対値よりも有意に低いことを示している。つまりバーコードの構造化によって変異を防ぐことができることを示している。

[0076] (異なる RNA ライブラリを区別するためのバーコード)

第 2 のライブラリを使用して、バーコードがすべてのバージョンの共通プール内の RNA ライブラリの異なるバージョンを区別するのに役立つかどうかを実験した。このため、第 2 のライブラリは、インビトロ転写の前にプライマー Pr__d 2 a、Pr__d 2 b、Pr__d 2 c、Pr__d 2 d を使用してバッチバーコード (第 2 のバーコード) を RNA に付与し 4 つの異なるバージョンに区別した。図 6 に示したように、RNA ライブラリの 4 つの異なるバージョンが、NAI 又は DMS で修飾されるか、又はそれぞれのコントロールとして取り扱われた。精製ステップの後、ライブラリの 4 つのバージョンの等量を混合することにより、プールされたサンプルが各処理条件に対し

て作成された。ライブラリの4つの異なるバージョンとプールされたサンプルのそれぞれは、連続したステップで同じ方法で処理された。

[0077] これらのデルタ変異率を、ID1の対象シーケンスをX軸としてプロットした図8(a)及び図8(b)に示す。デルタ変異率は、第1のライブラリ及び第2のライブラリの4つの群すべてを示した(データはプールされたサンプルからのデータである。)。NAI(図8(a))の場合、第1のライブラリと第2のライブラリのデルタ変異率はわずかに異なるが、いずれのライブラリについても非拘束ヌクレオチド領域では変異率が高く、構造プロービングが二次構造に関する情報を反映していることを示している。予測には、ViennaRNAパッケージを使用した。DMS(図8(b))の場合、ライブラリ間の違いはそれほど顕著ではないが、NAIと同様に、構造情報は、拘束されていないことが予測される領域のデルタ変異率がより高い値を示している。実験を行った条件ではDMSは塩基GおよびUの修飾効率が低いため、塩基CおよびAのみがより高い変異率を示す。プールされたサンプルの結果と個別に処理されたサンプルの結果を比較すると、グラフの曲線間に良好なオーバーラップが見られる(図8(c)及び図8(d))。したがって、プーリングは実験の結果に大きな影響を与えないと考えられ、バーコードとしての機能を果たしている。

[0078] (第2の設計によるRNAライブラリの2次構造情報)

図8は、単一IDの変異プロファイルのみを示している。次にすべてのIDの変異プロファイルを分析し、ViennaRNAパッケージで予測した二次構造と比較した。図9は、第2のライブラリをそれぞれ単独で又はプールしてNAI又はDMSで化学修飾したときの、塩基対を形成すると予測された領域(図9の黒い領域)と非結合であると予測された領域(図9の灰色の領域)のデルタ変異率の絶対値を別々にプロットしたバイオリンプロットである。図9(a)は、NAIで処理したサンプル、図9(b)は、DMSで処理したサンプルであり、それぞれのx軸に示したIDのうち、ID1~28は構造化バーコード配列を、ID29~37は非構造化バーコード配列

を含む。この結果は、4つの個別のサンプル（図9の「バイオリン」の左側）とプールされたサンプル（図9の「バイオリン」の右側）の分布が非常に似ていることも示している。DMSの場合、塩基AとCの位置のみが考慮される。

[0079] この結果は、各IDのサンプルについて、NAI（図9（a））又はDMS（図9（b））を使用すると、非結合領域のデルタ変異率の絶対値が高くなることを示しており、ライブラリ内の各シーケンスについて、二次構造情報を取得できることを示している。さらに、プールされたサンプルとプールされていないサンプルでは、平均値と分布の幅（標準偏差）に大きな違いは認められなかった。DMSで修飾した場合は、S/N比が向上し、修飾されたRNAと修飾されていないRNAの分布のオーバーラップは少なくなることを示している。一方で構造化バーコード（ID1～28）に対して非構造化バーコード（ID29～37）においては分布の顕著なオーバーラップが観察されることが多い。これは非構造化バーコードにおいてRNA構造予測によるデータと一致しないことを意味しており、非構造化バーコードが解析対象のRNAの構造に影響を与えたことを示している。

[0080] [実施例2] 構造化バーコードの使用によるバーコード識別の正確性

全体で54種類のRNA構造が混在するマルチプレックス化されたライブラリ（RNAプローブライブラリ）に対して、96種類の構造化バッチバーコードを用意した。その後マッピングのために、ライブラリに含まれる54種類すべてのRNA構造に異なるバーコードを付与し、 96×54 種類のリファレンスファイルを作成した。実際にそのうちIDが異なる2種類のバッチバーコードを付加したRNAプローブライブラリを試験管内合成し、DMSによる変異プロファイル実験を行った。検証実験のために異なる構造化バッチバーコードに対して対応したインデックスを付与し、次世代シーケンシング解析を行った。その後、得られたすべてのリードをリファレンスファイルにマッピングをした。本解析ではSTARアライナーソフトウェアを用いてマッピングした。その結果を図10及び図11に示す。

- [0081] 図10は構造化バッチバーコード1を用いた実験であり、横軸にシーケンスとマッピングにより実際に判定されたID、縦軸にリード数の合計（Depth_{sum}）を示す。構造化バッチバーコード1を用いた変異プロファイル反応系では修飾剤を用いておらず、RNA構造選択的な変異導入の効果がない。実際に構造化バッチバーコード1のほとんどはID1と正しい判定を受けている。18種類のIDにマッピングが間違っ行われたという判定がされているが、正解のID1に対して、その他のIDのリード数では1/1000~1/10000以下と非常に小さいため、変異プロファイルのデータの解釈に影響を与えない。
- [0082] 図11は構造化バッチバーコード2を用いた実験であり、横軸にシーケンスとマッピングにより実際に判定されたID、縦軸にリード数の合計（Depth_{sum}）を示す。構造化バッチバーコード2を用いた変異プロファイル反応系では修飾剤を用いており、RNAの高次構造選択的に変異導入が行われる。図10に対して図11では、変異導入のため一定以上のリード数の検出が行われたIDが増えた現象が確認されたものの、図10と同様に大多数のリード数はID2と正解の判定を受けた。加えて、正しいID（ID2と判定されたもの）に比して間違っ判定されたID（ID2以外と判定されたもの）のリード数の総和は1/100~10000以下と非常に小さいため、変異プロファイルのデータの解釈に影響を与えない。
- [0083] 実際にライブラリ中の54種類のRNAごとに正確性（正しいIDと判定された割合）を確認した（図12、13）。その結果、非修飾条件における正確性は平均99.91%、変異導入条件では平均99.44%であり、変異導入条件においても高い正確性を維持していた。
- [0084] 以上より、構造化バッチバーコードは変異プロファイルにおいて、マッピングの正確性を損なわずに正解のバーコードIDを他の間違っIDと明確に区別できるため、複数の異なる条件を同時に混ぜるマルチプレックス化に有用である。
- [0085] [実施例3]バーコードと他のバーコード（インデックス）との組み合わせに

よるマルチプレックス化の効果

RNAを用いた変異プロファイル反応を終え、DNAに変換したのちに市販のインデックスプライマー（例、Nextera XT Index Kit <イルミナ社>）などと組み合わせることで、サンプルの由来や条件の複雑性を上げることができる。図14は縦軸にイルミナ社の配列に基づいたインデックスプライマー（バーコードとして機能する）、横軸に実施例2で調製した構造化RNA ID7のサンプルをマッピングした際に判定されたIDを示す。カラースケールはリード数の平均値を示す。

[0086] 上記より、どのインデックスプライマーにおいても構造化バッチバーコード（ID）は高い正確性を保って識別可能であることがわかった。つまり、バッチバーコードに加え、複数の形態のDNAバーコードを組み合わせることで検体数を大規模に拡張可能と言える。例えば、10種類のインデックスプライマーと96種類の構造化バーコードを使用することで10×96の960通りの条件を設定できる。

[0087] [実施例4]構造化バーコードを用いた次世代シーケンス解析における直交性

全体で異なる1500種類のRNAプローブが混在するマルチプレックス化されたライブラリ（RNAプローブライブラリ）に対して、32種類の構造化バッチバーコードを用意した。その後マッピングのために、1500種類すべてのRNAに異なるバッチバーコードを付与し、32×1500種類（48000種類）のリファレンスファイルとともに実際にRNAプローブライブラリを試験管内合成した。次に、構造化バッチバーコードが付与されたRNAプローブライブラリ群を用いたプロファイル解析を行った。検証実験のために32の異なる構造化バッチバーコードに対してすべて32種類異なるインデックスプライマーを用いてインデックス（Index ID）を付与し、次世代シーケンサー（MiSeq<イルミナ社>）によるシーケンシング解析を行った。その後、インデックスにより32種類のファイルに分配した。バーコードが正しく機能すれば、インデックスID1に相当するファイルには構造化バッチバーコードID1が付与されたRNAプローブライ

ブラリが含まれる。その後、得られたすべてのリードをリファレンスファイルにマッピングをした。本解析ではSTARアライナーソフトウェアを用いてマッピングした。

[0088] 図15では、横軸に正しいインデックス (Index ID)、縦軸に実際にシーケンスとマッピングにより判定された構造化バッチバーコードID (Batch Barcode ID) を示す。ヒートマップの色はマッピングされたリード数のRNAプロブライブラリ内での平均値 (Depth_mean) を示す。図15に示す通り、すべての構造化バーコードについて、正しいIDへと割り振られることを確認した。さらに、図16に示す通り、誤判定はライブラリ中の1500種類RNAに対してほとんどが0または10種類未満に対して発生することがわかり、ライブラリ全体のRNAに対して影響はごく小さい。

[0089] 加えてこれらの誤判定されたRNA種類に対してのリード数は正しいIDに比しておよそ1/100~10000以下であるため、さらに影響が小さく、誤判定がプロファイルの結果への解釈に影響を与えないと言える (図17)。このため、構造化バッチバーコードは意図した通り高い直交性を有していると言え、バーコードとして機能したことを示す。なお、図16にて、一部800種類、130種類程度の混在が確認されたデータ点があるが、それは隣り合ったチューブ間で連続して発生していることやバーコードにおける類似性がないため、人為的ミスによるコンタミネーションと判断され、特定の構造化バーコードによる問題ではない。

[0090] なお、本実施例で用いた構造化バーコード配列の例 (ID12及びID28) を図18に示す。ID12の構造化バーコードRNAは、22塩基長：5'-GCUAGAAGAUUUGUCUUCUGGU-3' (配列番号9) で4塩基のループ構造を含む。一方、ID28の構造化バーコードRNAは、19塩基長：5'-UUGCGAGAUUUCUCGCGA-3' (配列番号10) で3塩基のループ構造を含む。このように、構造化バーコードは塩基配列のみならず長さや高次構造を変更することもできるため、その

組み合わせをさらに拡大することが可能である。

[0091] 以上のように、構造化バーコードは複数の反応条件での構造プロービング試験の多重化（マルチプレックス化）が可能である。応用としては異なる反応組成、実験環境条件を複数揃えた上で構造プロービング試験を行うことができ、それらの異なる条件がRNA構造に与える影響を大規模にスクリーニングすることができる。例えば、以下の参考文献[1]～[3]に例示される方法などを用いて、構造変化をきたすことが知られている分子や条件の評価を一度に複数種類行うスクリーニングへと拡張することができる。

[0092] 参考文献

[1] Komatsu, K. R., Taya, T., Matsumoto, S., Miyashita, E., Kashida, S., & Saito, H. (2020). RNA structure-wide discovery of functional interactions with multiplexed RNA motif library. *Nature communications*, 11(1), 1-14.

[2] Tapsin, S., Sun, M., Shen, Y., Zhang, H., Lim, X. N., Susanto, T. T., ... & Wan, Y. (2018). Genome-wide identification of natural RNA aptamers in prokaryotes and eukaryotes. *Nature communications*, 9(1), 1-10.

[3] Corley, M., Flynn, R. A., Lee, B., Blue, S. M., Chang, H. Y., & Yeo, G. W. (2020). Footprinting SHAPE-eCLIP Reveals Transcriptome-wide Hydrogen Bonds at RNA-Protein Interfaces. *Molecular Cell*, 80(5), 903-914.

請求の範囲

- [請求項1] RNAの高次構造を解析するための方法であって、
- (a) 解析対象RNAにバーコード配列を付加した1又は複数のRNAプローブを調製する工程、
- (b) 前記RNAプローブとRNA修飾剤とを接触させる工程、及び
- (c) 工程(b)で得られたRNAプローブの配列中で、修飾を受けた塩基の位置と頻度を検出する工程、を含み、前記バーコード配列は前記RNA修飾剤との反応が抑制される構造を有する、方法。
- [請求項2] 前記工程(c)が以下の工程：
- (c1) 工程(b)で得られたRNAプローブの混合物を鋳型として逆転写酵素により相補DNAを合成する工程、
- (c2) 前記相補DNAの塩基配列を決定し、前記バーコード配列を含む塩基配列を整列させる工程、及び
- (c3) 前記整列させた塩基配列に生じた変異の位置と頻度を検出する工程、
- を含む、請求項1に記載の方法。
- [請求項3] 前記RNA修飾剤が、前記RNAプローブ中の拘束ヌクレオチドを選択的に修飾するとき、前記バーコード配列が塩基対を形成しない配列である請求項1又は2に記載の方法。
- [請求項4] 前記RNA修飾剤が、前記RNAプローブ中の非拘束ヌクレオチドを選択的に修飾するとき、前記バーコード配列が複数の塩基対を含む構造を形成する請求項1又は2に記載の方法。
- [請求項5] 前記複数の塩基対を含む構造が、相補的な二本鎖構造、三重鎖構造又は四重鎖構造である請求項4に記載の方法。
- [請求項6] 前記複数の塩基対が、ステムループ構造又はシュードノット構造のステム部位に存在する請求項4又は5に記載の方法。
- [請求項7] 前記複数の塩基対を含む構造が、ステムループ構造であり、ステム部位に1つ以上のバルジ及び／又は内部ループ構造を有する請求項

4～6のいずれか一項に記載の方法。

[請求項8] 前記複数の塩基対を含む構造が、PDB (Protein Data Bank) に登録されているRNA構造又はその改変体である請求項4～7のいずれか一項に記載の方法。

[請求項9] 前記解析対象RNAが、少なくとも1つのRNAモチーフを含む、請求項1～8のいずれか一項に記載の方法。

[請求項10] 複数の塩基対を含む構造を形成するバーコード配列が付加された解析対象RNAを含むRNAプローブ。

[請求項11] 前記複数の塩基対を含む構造が、相補的な二本鎖構造、三重鎖構造又は四重鎖構造である請求項10に記載のRNAプローブ。

[請求項12] 前記複数の塩基対が、ステムループ構造又はシュードノット構造のステム部位に存在する請求項10又は11に記載のRNAプローブ。

[請求項13] 複数の塩基対を含む構造を形成するバーコード配列がそれぞれの解析対象RNAに付加された複数のRNAプローブを含むRNAプローブライブラリ。

[請求項14] 請求項13に記載のRNAプローブライブラリの2以上の複製物からなるRNAプローブライブラリ群であって、複製されたすべてのRNAプローブは、さらに第2のバーコード配列を含み、前記第2のバーコード配列は、1つのライブラリ内ではすべて同一配列であるが他のライブラリとの間では識別可能である、RNAプローブライブラリ群。